PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-143493
(43)Date of publication of application : 28.05.1999

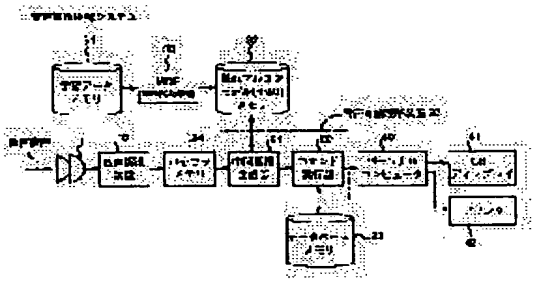
(51)Int.Cl. G10L 3/00
G10L 3/00
G10L 3/00
G06F 15/18
G06F 17/28
G06F 17/30
// G06F 3/16

(21)Application number : 09-303075
(22)Date of filing : 05.11.1997

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : MASATAKI HIROKAZU

(54) DEVICE AND SYSTEM FOR UNDERSTANDING VOICE WORD

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a device and a system for understanding the voice word capable of firmly and correctly understanding the voice, and executing the appropriately responding processing compared with the conventional technology.
SOLUTION: An MCE learning part 30 learns a hidden Markov model to convert sentence data into intermediate words corresponding thereto so that the discrimination error is minimized based on the learned data. An intermediate word generation part 21 converts the voice sentence of the result of the speech recognition which is speech recognized and includes the retrieval condition using a hidden Markov model to convert the sentence data into the intermediate words corresponding thereto. After a command execution part 22 coverts the generated intermediate words into the prescribed database words corresponding to the database, it retrieves the database based on the retrieval condition included in the database words, acquires the data to satisfy the retrieval condition, and executes the processing responding to the intermediate words on the data.



LEGAL STATUS

[Date of request for examination]	05.11.1997
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3088364
[Date of registration]	14.07.2000
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平11－143493

(43)公開日　平成11年(1999)５月28日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ１０Ｌ	3/00	５７１	Ｇ１０Ｌ	3/00	５７１Ｈ
		５３５			５３５
		５６１			５６１Ｇ
Ｇ０６Ｆ	15/18	５６０	Ｇ０６Ｆ	15/18	５６０Ｇ
	17/28			3/16	３２０Ｈ

審査請求　有　請求項の数２　ＯＬ　（全 14 頁）　最終頁に続く

(21)出願番号　特願平9－303075

(22)出願日　平成９年(1997)11月５日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　政瀧　浩和
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

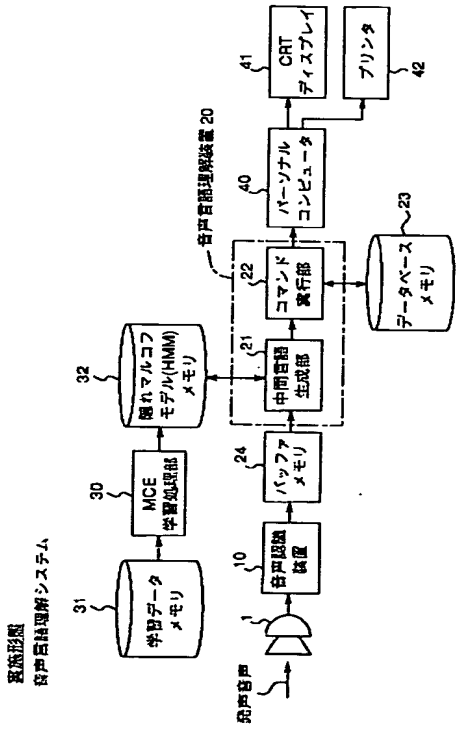
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　音声言語理解装置及び音声言語理解システム

(57)【要約】

【課題】　従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置及び、音声言語理解システムを提供する。

【解決手段】　ＭＣＥ学習処理部３０は学習データに基づいて識別誤りが最小となるように文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習する。中間言語生成部２１は、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを用いて、音声認識されかつ検索条件を含む音声認識結果の音声文を中間言語に変換して生成する。コマンド実行部２２は、生成された中間言語を、データベースに対応した所定のデータベース言語に変換した後、データベース言語に含まれる検索条件に基づいて、データベースを検索して、検索条件を満たすデータを獲得し、そのデータについて中間言語に対応した応答する処理を実行する。



【特許請求の範囲】

【請求項１】　発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であって、

検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、

複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、

上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、

上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする音声言語理解装置。

【請求項２】　発声音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解システムであって、

発声音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、

検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、

複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、

上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、

上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする音声言語理解システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置、並びに、音声認識装置及び音声言語理解装置とを備えた音声言語理解システムに関する。

【０００２】

【従来の技術及び発明が解決しようとする課題】近年、隠れマルコフモデルを用いた音響モデル、及びＮ－ｇｒａｍを用いた言語モデルを用いた連続音声認識が盛んに研究されており、数万語彙の認識でも、単語認識率が９０％以上とかなり実用レベルに近くなっている。しかしながら、音声認識技術を用いたアプリケーションを考えた場合、読み上げた文章をそのまま出力するディクテーションシステムを除くと、旅客機案内システム、電話番号案内システム、音声翻訳システム等、音声認識結果を理解し、ユーザーに情報を提供するいわゆる「音声理解システム」の方が応用分野が広いと考えられる（例えば、従来技術文献１「坂井信輔ほか，“音声入力を用いたパソコンネットワーク旅客機空席案内システムの試作”，電子情報通信学会技術報告，ＳＰ９４－８９，ｐｐ．２９－３６，１９９５年１月」参照。）。

【０００３】現在、音声理解システムのための言語理解の技術は、発話の内容を構文で限定したものや文法理論を用いたもの（例えば、従来技術文献２「Ｓ．Ｓｅｎｅｆｆ，“ＴＩＮＡ：Ａ Ｎａｔｕｒａｌ Ｌａｎｇｕａｇｅ Ｓｙｓｔｅｍ ｆｏｒＳｐｏｋｅｎ Ｌａｎｇｕａｇｅ Ａｐｐｌｉｃａｔｉｏｎｓ”，Ｃｏｍｐｕｔａｔｉｏｎａｌ Ｌｉｎｇｕｉｓｔｉｃｓ，Ｖｏｌ．１８，Ｎｏ．１，１９９２年３月」参照。）が主流である。発話内容を構文で限定する手法は、理解率は高いと考えられるが、計算機が受理できる発話内容をユーザーが事前に知っていなければならず、ユーザーへの負担が大きく、使いやすいシステムとは言えない。

【０００４】また、文法理論を用いた方法は、構文による手法よりは、発話内容の自由度が高いが、文法的に正しい文章でないと、理解できないという問題がある。しかしながら、音声認識で広く使われているＮ－ｇｒａｍ言語モデルは、認識率の観点からは非常に有利とされているが、直前の（Ｎ－１）単語から次の単語への接続を確率で表現するという極めて単純なモデルであるため、局所的な制約しか表現できず、文全体として必ずしも文

法的に正しい文章を出力するとは限らない。従って、認識結果に誤りが含まれる場合、正しく言語理解を行うのは困難であると考えられる。また、実際のシステムの使用時には、ユーザーが文法的に正しい文章を発声するとは限らず、自然発話を理解するのは困難である。

【０００５】この問題を解決するため、認識結果文を言語理解部が受理できる部分に分割する手法等（例えば、従来技術文献３「Ｙ．Ｗａｋｉｔａ ｅｔ ａｌ．，“Ｃｏｒｒｅｃｔ ｐａｒｔｓ ｅｘｔｒａｃｔｉｏｎ ｆｒｏｍ ｓｐｅｅｃｈ ｒｅｃｏｇｎｉｔｉｏｎ ｒｅｓｕｌｔｓ ｕｓｉｎｇ ｓｅｍａｎｔｉｃ ｄｉｓｔａｎｃｅ ｃａｌｃｕｌａｔｉｏｎ，ａｎｄ ｉｔｓ ａｐｐｌｉｃａｔｉｏｎ ｔｏ ｓｐｅｅｃｈ ｔｒａｎｓｌａｔｉｏｎ”，ＡＣＬ，１９９７年」参照。）が提案されているが、分割を行うことにより、文章の大局的な情報を得ることができないと考えられる。

【０００６】また、従来技術文献４「遠藤充ほか，“音声による文例検索システムの検討”，日本音響学会講演論文集，２－Ｑ－１２，ｐｐ．１６３－１６４，１９９７年３月」においては、キーワードによる方法が提案されているが、キーワードのみでは文章の意味を正しく理解することができず、また、ユーザーインターフェース等でキーワードの間を補う必要がある。

【０００７】本発明の目的は以上の問題点を解決し、上記従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置及び、音声言語理解システムを提供することにある。

【０００８】

【課題を解決するための手段】本発明に係る請求項１記載の音声認識装置は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であって、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする。

【０００９】また、本発明に係る請求項２記載の音声言語理解システムは、発声音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解システムであって、発声音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備えたことを特徴とする。

【００１０】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【００１１】図１は、本発明に係る一実施形態である音声言語理解装置２０を備えた音声言語理解システムの構成を示すブロック図である。本発明に係る実施形態においては、統計的処理に基づく中間言語への変換を行うことにより、ユーザーの発声に対して頑健な理解が行える音声言語理解装置２０を提供することを特徴としている。

【００１２】ここで、音声言語理解装置２０は、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であり、（ａ）学習データメモリ３１に格納され、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデル（ＨＭＭ）をＭＣＥ学習処理部３０によって学習して得られた隠れマルコフモデルを記憶する隠れマルコフモデルメモリ３２と、（ｂ）複数の

項目名に対応したデータを含むデータベースを記憶するデータベースメモリ２３と、（ｃ）隠れマルコフモデルメモリ３２に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する中間言語生成部２１と、（ｄ）中間言語生成部２１によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、データベースメモリ２３に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理、具体的には表示処理を実行するコマンド実行部２２とを備えたことを特徴としている。

【００１３】まず、本実施形態の音声言語理解システムの概要について説明する。本実施形態では、音声言語理解システムとして、音声入力による指示により、データベースメモリ２３内データベースへアクセスし、ユーザーの要求する情報を表示するシステムを開示する。好ましい実施形態として構築したシステムは、スキー場案内システムであって、音声により、スキー場のデータの入ったデータベースにアクセスし、必要な情報を得るシステムである。システム全体の構成を図１に示す。本システムは、主に「音声認識装置１０」と「音声言語理解装置２０」とで構成される。

【００１４】音声認識装置１０では、入力された発声音声の波形データに対して特徴量計算を行った後、公知の隠れマルコフ網による音響モデル、及び公知の可変長Ｎ－ｇｒａｍによる言語モデルを用いて、単語グラフサーチ法により解の探索を行い、認識結果を出力する。音声認識装置１０の認識結果は音声言語理解装置２０に渡される。音声言語理解装置２０では、音声認識結果の単語列を、中間言語に変換し、中間言語の内容に基づいてデータベース言語の生成し、データベースから情報の検索を行い、中間言語の要求に応じて検索結果の表示を行う。本システムは、次の３つの動作を行うことができる。

（ａ）各スキー場のデータ（県・標高差・リフト数等１２項目）の表示（ＳＨＯＷＶＡＬＵＥ）、（ｂ）ユーザーが要求する条件を満たすスキー場の検索（ＳＨＯＷＬＩＳＴ）、及び（ｃ）スキー場の地図の表示（ＳＨＯＷＩＭＡＧＥ）。

【００１５】次いで、音声言語理解装置２０について説明する。その動作の概要を表１に示す。また、データベースメモリ２３内のデータベースの一例を表２に示す。

【００１６】

【表１】

入力文：“八方尾根スキー場の標高差を教えて下さい”
１．中間言語生成 “Ｒ＿ＳＨＯＷＶＡＬＵＥ　Ｏ＿標高差　Ｄ＿スキー場名　Ｃ＿＝　Ｖ＿八方尾根”
２．データベース言語への変換 “ＳＥＬＥＣＴ標高差　ＦＲＯＭスキー場データ ＷＨＥＲＥスキー場名＝八方尾根”

【００１７】表１の２．においてデータベース言語の一例について示している。この例では、「スキー場データ」というデータベース名のデータベースから（ＦＲＯＭ）、「スキー場名＝八方尾根」という検索条件のもと（ＷＨＥＲＥ）で、項目名「標高差」のデータを検索（ＳＥＬＥＣＴ）しなさいというデータベース言語である。

【００１８】

【表２】

データベースの一例
スキー場データ

スキー場名	県	標高差	入場者数	
志賀高原	長野	５００	１０００００	
野沢温泉	長野	１１００	９００００	
妙高赤倉	新潟	８００	８００００	
八方尾根	長野	１０００	７００００	←動作例
栂池高原	長野	７００	６００００	

（注）動作例
３．スキー場名＝八方尾根の行を検索
４．標高差を出力
【００１９】音声言語理解装置２０の処理は、音声認識結果をデータベースアクセス用の中間言語に変換することにより行う。本システムで用いた中間言語は次の要素から構成される。
（ａ）Ｒ＿（コマンド名）
要求動作の指定（Ｒｅｑｕｅｓｔ）
（ｂ）Ｏ＿（対象物名）
動作の対象（Ｏｂｊｅｃｔ）
（ｃ）Ｄ＿（ドメイン名）
データベースの検索項目（Ｄｏｍａｉｎ）
（ｄ）Ｃ＿（比較方法）
データベース検索時の比較方法（Ｃｏｍｐａｒｉｓｏｎ）
（ｅ）Ｖ＿（値）
データベース検索時の比較値（Ｖａｌｕｅ）
【００２０】中間言語は、これらの要素の列として表現され、次の表で示すフォーマットで与えられる。
【００２１】
【表３】

Ｒ＿（コマンド名）	Ｏ＿（対象物名１）…	Ｏ＿（対象物名ｍ）
Ｄ＿（ドメイン名１）	Ｃ＿（比較方法１）	Ｖ＿（値１）
…		
Ｄ＿（ドメイン名ｎ）	Ｃ＿（比較方法ｎ）	Ｖ＿（値ｎ）

【００２２】以下に、自然言語から中間言語への変換例を挙げる。以下の変換例は、図１の学習データメモリ３１に、学習データの文章データとそれに対応する中間言語データとして格納される。
（ａ）八方尾根スキー場の標高差はいくらですか。
→Ｒ＿ＳＨＯＷＶＡＬＵＥ　Ｏ＿標高差
Ｄ＿スキー場名　Ｃ＿＝　Ｖ＿八方尾根
（ｂ）標高差が１０００ｍ以上のスキー場を教えて下さい。
→Ｒ＿ＳＨＯＷＬＩＳＴ　Ｏ＿スキー場名　Ｄ＿標高差　Ｃ＿＞＝　Ｖ＿１０００
（ｃ）八方尾根のゲレンデマップを見せて下さい。
→Ｒ＿ＳＨＯＷＩＭＡＧＥ　Ｏ＿ゲレンデマップ
Ｄ＿スキー場名　Ｃ＿＝　Ｖ＿八方尾根
【００２３】音声言語理解装置２０のの一連の動作を表１及び表２に示す。音声言語理解装置２０は、音声認識結果が入力されると、次の順序で処理を行う。
（１）音声認識結果から中間言語への変換処理、（２）中間言語の、対象物名、ドメイン名（表２における項目名である。）、比較方法、及び比較値からデータベース言語を生成する処理、（３）条件に適合するデータをデータベースから検索し、動作の対象情報を獲得する処理、及び、（４）対象物名に対して中間言語のコマンド名で規定された動作を実行する処理。ここで、上記（１）の処理は図１の中間言語生成部２１によって実行され、上記（２）、（３）及び（４）の処理は図１のコマンド実行部２２によって実行される。なお、データベース言語は公知のＳＱＬ言語と類似した言語を用いており、中間言語は、データベース言語へ必ず正しく変換されるように設計されている。
【００２４】次いで、自然言語から中間言語への変換について説明する。音声言語理解装置１０において、最も重要かつ困難な部分は、音声認識の入力文章から中間言語への変換部分である。本実施形態は、これを統計的手法に基づいて実行する方法を用いる。
【００２５】統計的手法による自然言語から中間言語への変換では、単語系列Ｗが与えられたとき、次式を満たす中間言語列Ｓを求めることにより、最適な中間言語を得ることができる。
【数１】

$$\underset{S}{\mathrm{argmax}} P(S \mid W)$$

ここで、Ｐ（Ｓ｜Ｗ）は、単語系列Ｗが与えられたときの中間言語列Ｓを取り得る確率であり、数１は、中間言語列Ｓを変化したときに確率Ｐ（Ｓ｜Ｗ）が最大となるときの中間言語列Ｓを表わす。
【００２６】上記数１は、ベイズ則を用いると、次式のように表される。
【数２】

$$\underset{S}{\mathrm{argmax}} P(S \mid W) = \underset{S}{\mathrm{argmax}} P(W \mid S)\, P(S)$$

上記数２で、確率Ｐ（Ｗ｜Ｓ）は、中間言語から音声認識結果が出力される確率を意味する。この確率を直接的に求めるのは困難なため、次式の近似を考える。
【数３】

$$P(W \mid S) \fallingdotseq \Sigma P(W \mid s_i)$$

【００２７】すなわち、中間言語のそれぞれの要素は独立しており、また、中間言語のそれぞれの要素は、特定の単語のみを出力するのではなく、入力文の全ての単語をある確率で出力すると仮定する。この仮定により、誤

認識が生じた際や、不完全な文が入力された場合でも、中間言語への変換が容易になり、頑健な言語理解が可能になる。この確率Ｐ（Ｗ｜Ｓ）を求めるモデルを文生成モデルと呼ぶ。一方、確率Ｐ（Ｓ）は、中間言語の事前確率で、入力文とは全く独立に求められる。統計的手法による自然言語から中間言語への変換の精度は、これらの確率の精度に依存する。

【００２８】次いで、これらの確率を求めるための文生成モデルについて述べる。上記確率Ｐ（Ｗ｜s_i）、即ち、中間言語の各要素s_iから音声認識結果文を与えるモデルとして、隠れマルコフモデル（ＨＭＭ）を用いる（図４参照）。隠れマルコフモデルは、図４に示すように、複数の状態から構成され、単語が入力される毎に、状態ｉから状態ｊへ確率a_{ij}で遷移し、遷移後の状態ｊから確率b_{j,w_t}で単語を出力するモデルである。隠れマルコフモデルは、音声認識の音響モデルにおいては、通常Ｌｅｆｔ　ｔｏ　Ｒｉｇｈｔ型のモデルが用いられるが、ここで用いるモデルは、全ての状態間遷移が可能なエルゴディックモデルを考える。ここで、用いる隠れマルコフモデル（ＨＭＭ）のパラメータは次の通りである。

（ａ）状態数：Ｍ

（ｂ）状態ｉから状態ｊへの遷移確率：a_{ij}

（ｃ）状態ｊから単語w_tへの出力確率：b_{j,w_t}

【００２９】隠れマルコフモデルのパラメータの推定は、文章とそれに対応する中間言語列のデータを大量に容易し、Ｐ（Ｗ｜s_i）の値が最大になるように決定する（最尤推定法による）。隠れマルコフモデルの場合、観測データに対応する状態系列が分からないため、公知のＥＭ（Ｅｘｐｅｃｔａｔｉｏｎ－Ｍａｘｉｍｕｍ）アルゴリズムによって最尤推定を行う。隠れマルコフモデルの場合は特に、バーム・ウェルチ（Ｂａｕｍ－Ｗｅｌｃｈ）アルゴリズムと呼ばれる。隠れマルコフモデルは、中間言語の各要素毎に作成し、文が入力されると、全てのモデルが独立に、平行して動作する。隠れマルコフモデルを用いて、入力文から中間言語への変換を行う際は、公知のビタビ（Ｖｉｔｅｒｂｉ）アルゴリズムを用いてそれぞれの要素に対して文の生成確率のみを求め、Ｒ＿，Ｏ＿，Ｄ＿，Ｃ＿，Ｖ＿のそれぞれのグループ内で最も確率の高い要素を選び、中間言語列を得る（図５参照。）。

【００３０】次いで、中間言語への変換の高精度化のための、図１のＭＣＥ学習処理部３０によって実行される公知のＭＣＥ（Ｍｉｎｉｍｕｍ　Ｃｌａｓｓｉｆｉｃａｔｉｏｎ　Ｅｒｒｏｒ；識別誤り最小法）トレーニングによる学習について説明する。通常、隠れマルコフモデルの学習は、公知のＥＭアルゴリズムによって行われる。ＥＭアルゴリズムは、基本的には最尤推定法であり、本実施形態で用いる文生成モデルに使用した場合、不都合が生じる場合がある。例として、次の２つの場合について考える。（１）長野県のスキー場を教えて下さい。

→Ｒ＿ＳＨＯＷＬＩＳＴ　Ｏ＿スキー場名　Ｄ＿県　Ｃ＿＝　Ｖ＿長野

（２）長野県以外のスキー場を教えて下さい。

→Ｒ＿ＳＨＯＷＬＩＳＴ　Ｏ＿スキー場名　Ｄ＿県　Ｃ＿＜＞　Ｖ＿長野

【００３１】この場合、文（１）と文（２）との差は、単語「以外」があるかないかのみの違いであり、その差が中間言語の「Ｃ＿＝」と「Ｃ＿＜＞」との差となる。しかしながら、上述の文生成モデルであると、「Ｃ＿＜＞」のモデルでは、文（１）例に出現する単語全てに対して比較的高い確率を出力し、（１）の文が入力された場合「Ｃ＿＝」と「Ｃ＿＜＞」との区別が困難になる可能性がある。このため、類似した文に対する識別度を向上させるため、隠れマルコフモデルに対して、ＭＣＥ学習を行う。

【００３２】ＭＣＥに基づく学習法は、クラスの識別に用いる尺度を識別関数g_kとし、あるサンプルＸに対する識別関数の差で表される識別誤り関数$d_k(X, \Lambda)$から、シグモイド（ｓｉｇｍｏｉｄ）関数で現れる損失関数ｌ（d_k）を用いて実効的な識別誤り数を評価し、この識別誤り数を最小化する基準でモデルパラメータΛを求める方法である。ここで、用いる識別関数、識別誤り関数、及び損失関数を次式に示す。

【００３３】（ａ）識別関数

【数４】$g(X, \Lambda) = \log[L(\chi)]$

ここで、Ｌ（χ）は、隠れマルコフモデルによる文生成確率Ｐ（Ｗ｜s_i）を表わす。（ｂ）識別誤り関数

【数５】

$$d = -g_k(X, \Lambda) + \log\left(\left\{1/(K-1)\right\} \sum_{j, j \neq k} \exp\left[\eta g_l(X, \Lambda)\right]\right)^{1/\eta}$$

ここで、ｋは読み込んだデータの中間言語に含まれる要素で、ｊはｋのグループに属する中間言語の要素である。（ｃ）損失関数

【数６】

$$l(d_k) = 1/(1+\exp[-a(d_k+b)])$$

ここで、ａ及びｂは予め経験的に決定される定数である。

【００３４】損失関数ｌ（d_k）に対して、最急降下法を用いてて漸化的にパラメータΛを調整しながら、最適パラメータを求める。

【数７】$\Lambda h_{t+1} = \Lambda h_t - \varepsilon \nabla l(d_k(X; \Lambda h)$

【００３５】図６は、図１のＭＣＥ学習処理部３０によって実行されるＭＣＥ学習処理を示すフローチャートである。ＭＣＥ学習処理部３０は、学習データメモリ３１

内の学習データの文章データとそれに対応する中間言語データに基づいて、以下のＭＣＥ学習処理を実行することにより、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して隠れマルコフモデルメモリ３２に格納する。

【００３６】図６において、まず、ステップＳ０でパラメータｉに１を代入し、ステップＳ１で学習データメモリ３１から、文章データ及び中間言語データをそれぞれ１文読み込む。次いで、ステップＳ２で、数６を用いて、損失ｌを計算し、ステップＳ３で数７を用いて隠れマルコフモデル（ＨＭＭ）の各パラメータを更新する。そして、ステップＳ４で処理すべき学習データがまだあるか否かが判断され、ＹＥＳのときはステップＳ１に戻り上記の処理を繰り返す一方、ステップＳ５では全ての文について総損失L_iを計算する。そして、ステップＳ６では、学習終了判定が実行されて、$|L_i-L_{i-1}|<C$（ここで、Ｃは定数である。）であるか否かが判断され、ＮＯのときは終了していないとして、ステップＳ７でパラメータｉを１だけインクリメントしてステップＳ１に戻り上記の処理を繰り返す。ステップＳ６でＹＥＳのときは、終了したと判断して当該ＭＣＥ学習処理を終了する。ここで、学習データメモリ３１内の学習データの一例を次の表に示す。

【００３７】

【表４】

文章データ ／ 中間言語データ
八方尾根スキー場の標高差を教えて下さい。 ／ R__SHOWVALUE O__標高差 D__スキー場名 C__= V__八方尾根
標高差が１０００メートル以上のスキー場を教えて下さい。 ／ R__SHOWLIST O__スキー場名 D__標高差 C__>= V__１０００

【００３８】図７は、図１の中間言語生成部２１によって実行される中間言語生成処理を示すフローチャートである。図７において、まず、ステップＳ１１でバッファメモリ２４から音声認識結果を１文読み込む。次いで、ステップＳ１２で音声認識結果の単語列Ｗ（Ｌ単語）に対して、次式を用いて中間言語の各要素s_i（$1 \leqq i \leqq N$；Ｎは中間言語の要素数である。）毎に隠れマルコフモデル（ＨＭＭ）による文生成確率Ｐ（Ｗ｜s_i）を計算する。

【数８】

$$P(W|s_i) = \underset{x,\,y}{\mathrm{argmax}} \prod_{l=1}^{L} a_{xy}(s_i) \cdot b_{y,w_l}(s_i)$$

ただし、w_lは単語列Ｗのｌ番目の単語$a_{xy}(s_i)$は中間言語の要素s_iのモデルにおける、状態ｘから状態ｙへの遷移確率$a_{xy}(s_i)$は中間言語の要素s_iのモデルにおける、状態ｙから単語w_lへの出力確率を表す。

【００３９】次いで、ステップＳ１３で中間言語の各要素の文生成確率Ｐ（Ｗ｜s_i）に対して、各グループ内で尤度最大のものを選択する。すなわち、中間言語の各グループ（Ｒ__*，Ｏ__*，Ｄ__*，Ｃ__*，Ｖ__*）において、そのグループに属する中間言語要素の内、ステップＳ１２で求めた文生成確率Ｐ（Ｗ｜s_i）の最も高いものを選択する。次いで、ステップＳ１４で選択された中間言語の要素を所定のフォーマットにしたがって並べた後出力する。すなわち、ステップＳ１３で入力された中間言語の各要素を中間言語文のフォーマット、すなわち、Ｒ__*，Ｏ__*，Ｄ__*，Ｃ__*，Ｖ__*の順番に従って並べ、中間言語を生成してコマンド実行部２２に出力する。さらに、ステップＳ１５で処理すべき音声認識結果がまだあるか否かが判断され、ＹＥＳのときはステップＳ１１に戻り上記の処理を繰り返す。一方、ＮＯのときは当該中間言語生成処理を終了する。

【００４０】図８は、図１のコマンド実行部２２によって実行されるコマンド実行処理を示すフローチャートである。図８において、まず、ステップＳ２１で中間言語生成部２１から中間言語を１文入力する。次いで、ステップＳ２２で、中間言語をデータベース言語（ＳＱＬ言語）に変換する。すなわち、この変換は、次の表に示すように機械的に行われる。

【００４１】

【表５】

中間言語：
“R__ＳＨＯＷＶＡＬＵＥ　Ｏ__標高差　Ｄ__スキー場名　Ｃ__＝
Ｖ__八方尾根”

データベース言語：
“ＳＥＬＥＣＴ標高差　ＦＲＯＭスキー場データ
　　ＷＨＥＲＥスキー場名＝八方尾根”

【００４２】ここで、データベース言語は、「ＳＬＥＥＣＴ（１）ＦＲＯＭスキー場データＷＨＥＲＥ（２）（３）（４）」の形式をしており、（１）から（４）までの項目を、それぞれ、中間言語のＯ__，Ｄ__，Ｃ__，Ｖ__等の頭文字を削除した物を並べることにより用意に変換が可能である。

【００４３】次いで、ステップＳ２３で変換されたデータベース言語に基づいて、所定の条件に適合するデータをデータベースメモリ２３から検索する。さらに、ステップＳ２４では、データベースより得られたデータを中間言語のコマンド名に応じて加工して出力する。すなわち、ステップＳ２３で得られた値を中間言語のコマンド名に応じて表示する。具体的には、次の表に示す表示処理を行う。

【００４４】

【表６】

コマンド名　　→　　表示内容
R__ＳＨＯＷＶＡＬＵＥ→データベースから得られた値を表示
R__ＳＨＯＷＬＩＳＴ→データベースからスキー場名の一覧を表示
R__ＳＨＯＷＩＭＡＧＥ→データベースから得られたファイル名の画像を表示

【００４５】ステップＳ２４における表示処理においては、表示内容のデータがコマンド実行部２２からパーソナルコンピュータ４０を介してＣＲＴディスプレイ４１に出力されて表示され、また、表示内容のデータがプリンタ４２に出力されて印字される。さらに、ステップＳ２５では、処理すべき中間言語があるか否かが判断され、ＹＥＳのときはステップＳ２１に戻り上記の処理を繰り返す一方、ＮＯのときは当該コマンド実行処理を終了する。

【００４６】図１において、ＭＣＥ学習処理部３０、中間言語生成部２１、及びコマンド実行部２２は例えばデジタル計算機で構成され、学習データメモリ３１、隠れマルコフモデル（ＨＭＭ）メモリ３２、バッファメモリ２４及びデータベースメモリ２３は例えばハードディスクメモリなどの記憶装置で構成される。

【００４７】図２に本実施形態で用いる連続音声認識装置１０のブロック図を示す。本実施形態の連続音声認識装置１０は、公知のワンーパス・ビタビ復号化法を用いて、入力される発声音声文の音声信号の特徴パラメータに基づいて上記発声音声文の単語仮説を検出し尤度を計算して出力する単語照合部４を備えた連続音声認識装置において、単語照合部４からバッファメモリ５を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、統計的言語モデル１３を参照して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う単語仮説絞込部６を備える。

【００４８】ここで用いる統計的言語モデル１３は、学習用テキストデータに基づいて言語モデル生成部（図示せず。）により生成されたものであって、統計的言語モデル１３は、例えば特開平９－１３４１９２号公報において開示されたように、品詞クラス間のバイグラム（Ｎ＝２）を基本としたものであるが、単独で信頼できる単語は品詞クラスより分離させ、単独のクラスとして取り扱い、さらに、予測精度を向上させるため、頻出単語列に関してはそれらの単語を結合して一つのクラスとして取り扱い、長い単語連鎖の表現を可能にさせ、こうして、生成されたモデルは、品詞バイグラムと可変長単語Ｎ－グラムとの特徴を併せ持つ統計的言語モデルとなり、遷移確率の精度と信頼性とのバランスをとられたものである。

【００４９】図２において、単語照合部４に接続され、例えばハードディスクメモリに格納される音素ＨＭＭ１１は、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号
（ｂ）受理可能なコンテキストクラス
（ｃ）先行状態、及び後続状態のリスト
（ｄ）出力確率密度分布のパラメータ
（ｅ）自己遷移確率及び後続状態への遷移確率

なお、本実施形態において用いる音素ＨＭＭ１１は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合ＨＭＭを変換して生成する。ここ

で、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布である。また、単語照合部４に接続され、例えばハードディスクに格納される単語辞書１２は、音素ＨＭＭ１１の各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【００５０】図２において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して単語照合部４に入力される。単語照合部４は、ワンーパス・ビタビ復号化法を用いて、バッファメモリ３を介して入力される特徴パラメータのデータに基づいて、音素ＨＭＭ１１と単語辞書１２とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部４は、各時刻の各ＨＭＭの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素ＨＭＭ１１及び単語辞書１２とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部４は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ５を介して単語仮説絞込部６に出力する。

【００５１】単語仮説絞込部６は、単語照合部４からバッファメモリ５を介して出力される単語仮説に基づいて、統計的言語モデル１３を参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として、バッファメモリ２４を介して音声言語理解装置２０内の中間言語生成部２１に出力し、上述の音声言語理解装置２０の処理が実行される。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びをいう。

【００５２】例えば、図３に示すように、（ｉ－１）番目の単語W_{i-1}の次に、音素列a_1，a_2，…，a_nからなるｉ番目の単語W_iがくるときに、単語W_{i-1}の単語仮説として６つの仮説Ｗａ，Ｗｂ，Ｗｃ，Ｗｄ，Ｗｅ，Ｗｆが存在している。ここで、前者３つの単語仮説Ｗａ，Ｗｂ，Ｗｃの最終音素は／ｘ／であるとし、後者３つの単語仮説Ｗｄ，Ｗｅ，Ｗｆの最終音素は／ｙ／であるとする。終了時刻t_eと先頭音素環境が等しい仮説（図３では先頭音素環境が“ｘ／a_1／a_2”である上から３つの単語仮説）のうち総尤度が最も高い仮説（例えば、図３において１番上の仮説）以外を削除する。なお、上から４番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素がｘではなくｙであるので、上から４番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に１つのみ仮説を残す。図２の例では、最終音素／ｘ／に対して１つの仮説を残し、最終音素／ｙ／に対して１つの仮説を残す。

【００５３】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも１つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【００５４】図２において、特徴抽出部２と、単語照合部４と、単語仮説絞込部６と、言語モデル生成部２０とは、例えば、デジタル電子計算機で構成され、バッファメモリ３，５は例えばハードデイスクメモリなどの記憶装置で構成され、音素ＨＭＭ１１と単語辞書１２と統計的言語モデル１３とは、例えばハードデイスクメモリなどの記憶装置に記憶される。

【００５５】以上実施形態においては、単語照合部４と単語仮説絞込部６とを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素ＨＭＭ１１を参照する音素照合部と、例えばＯｎｅ　Ｐａｓｓ　ＤＰアルゴリズムを用いて統計的言語モデル１３を参照して単語の音声認識を行う音声認識部とで構成してもよい。

【００５６】

【実施例】本発明者は、音声言語理解装置２０における言語理解率を評価するために、まず、正解文からの言語理解率を評価した。実験に用いたデータは、本特許出願人が所有するスキー場案内システムのために収集している会話で、現在、４４３文、７，５６９単語あり、語彙は２８１語である。全ての文章に、それに対応する中間言語を人手で作成している。言語理解のためのモデルは、最尤推定による隠れマルコフモデル（ＭＬ－ＨＭＭ）、及び、最尤推定後にＭＣＥ学習を行ったモデル（ＭＣＥ－ＨＭＭ）の２種類用意した。ただし、隠れマルコフ状態数は、いずれのモデルも２とした。

【００５７】評価は言語理解率で行った。ただし、言語理解率は、入力文章から中間言語へ正確に変換できた割合であり、中間言語の全ての要素が正しく変換できた場合のみ正解とする。最尤推定による隠れマルコフモデル（ＨＭＭ）を用いた場合、言語理解率は９６．０％とかなり高い率を得た。さらにＭＣＥ学習を行うことにより

言語理解率は９９．６と極めて高い率を得ることができた。

【００５８】以上説明したように、本実施形態によれば、隠れマルコフモデルを用いた統計的手法により、自然言語から中間言語への変換を行い、言語理解を行う音声言語理解システムを構築して、最尤推定による隠れマルコフモデルにおける言語理解率が９６．０％であり、さらにＭＣＥ学習を行うことにより、９９．６％と非常に高い確率で言語理解率が得られることを確認した。音声言語理解装置２０は、統計的手法を用いて処理するため、文法ルールの作成やキーワードの選択等の作業を必要とせず、また、モデルの学習には数千語程度のデータで良好な結果を得るため、短時間でのシステム構築が可能であるという利点がある。すなわち、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置２０及び、音声言語理解システムを提供することができる。

【００５９】

【発明の効果】以上詳述したように、本発明に係る請求項１記載の音声認識装置によれば、発声音声から音声認識装置によって音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解装置であって、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備える。従って、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解装置を提供することができる。

【００６０】また、本発明に係る請求項２記載の音声言語理解システムによれば、発声音声を音声認識して、音声認識された音声認識結果の音声文に基づいて、音声文の意味する検索条件の内容を理解して、データベースを参照して応答する処理を実行するための音声言語理解システムであって、発声音声を音声認識して、音声認識された音声認識結果の音声文を出力する音声認識装置と、検索条件を含む文章データと、それに対応しかつ少なくとも応答する処理の内容及びデータの項目名を含む所定の中間言語データとの対である学習データに基づいて、識別誤りが最小となるように、文章データをそれに対応する中間言語に変換するための隠れマルコフモデルを学習して得られた隠れマルコフモデルを記憶する第１の記憶装置と、複数の項目名に対応したデータを含むデータベースを記憶する第２の記憶装置と、上記第１の記憶装置に記憶された隠れマルコフモデルを用いて、上記音声認識装置によって音声認識されかつ検索条件を含む音声認識結果の音声文を上記中間言語に変換して生成する生成手段と、上記生成手段によって生成された中間言語を、上記データベースに対応した所定のデータベース言語に変換した後、上記データベース言語に含まれる検索条件に基づいて、上記第２の記憶装置に記憶されたデータベースを検索して、上記検索条件を満たすデータを獲得し、そのデータについて上記中間言語に対応した応答する処理を実行する実行手段とを備える。従って、従来技術に比較して頑健にかつ正確に音声理解を行うことができ、適切に応答する処理を実行することができる音声言語理解システムを提供することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である音声言語理解装置２０を備えた音声言語理解システムの構成を示すブロック図である。

【図２】　図１の音声認識装置１０の構成を示すブロック図である。

【図３】　図２の音声認識装置における単語仮説絞込部６の処理を示すタイミングチャートである。

【図４】　図１の隠れマルコフモデル（ＨＭＭ）メモリ３２に格納されたＨＭＭを示す状態遷移図である。

【図５】　図１の中間言語生成部２１の処理を示す説明図である。

【図６】　図１のＭＣＥ学習処理部３０によって実行されるＭＣＥ学習処理を示すフローチャートである。

【図７】　図１の中間言語生成部２１によって実行される中間言語生成処理を示すフローチャートである。

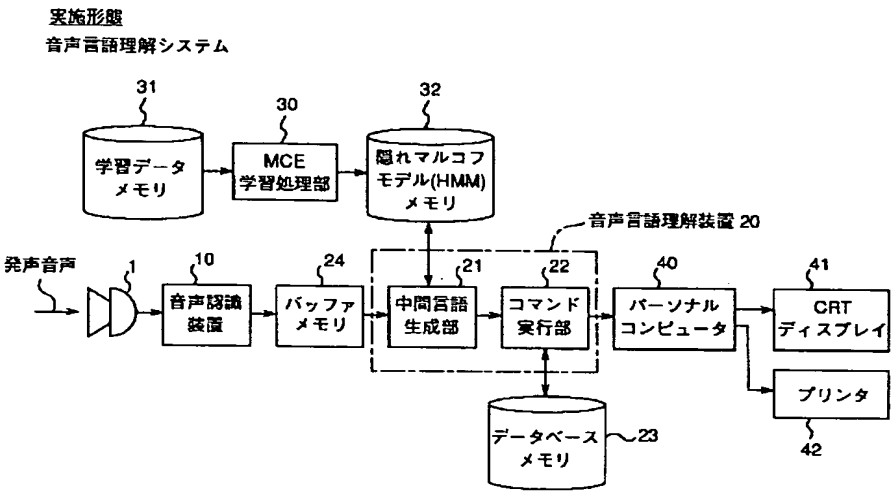
【図８】　図１のコマンド実行部２２によって実行されるコマンド実行処理を示すフローチャートである。

【符号の説明】

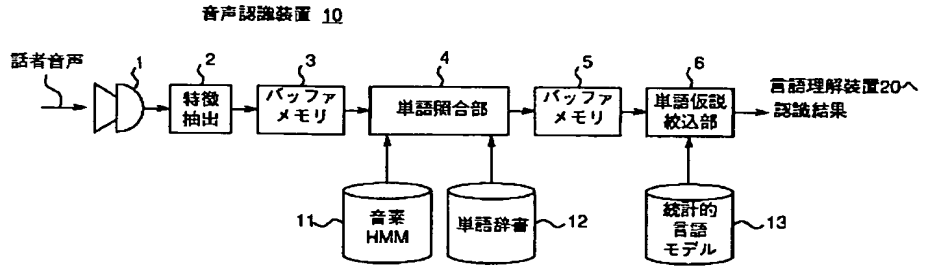
１…マイクロホン、
２…特徴抽出部、
３，５…バッファメモリ、
４…単語照合部、
６…単語仮説絞込部、
１１…音素ＨＭＭ、
１２…単語辞書、
１３…統計的言語モデル、

２０…音声言語理解装置、
２１…中間言語生成部、
２３…データベースメモリ、
２４…バッファメモリ、
３０…ＭＣＥ学習処理部、
３１…学習データメモリ、
３２…隠れマルコフモデル（ＨＭＭ）メモリ、
４０…パーソナルコンピュータ、
４１…ＣＲＴディスプレイ、
４２…プリンタ。

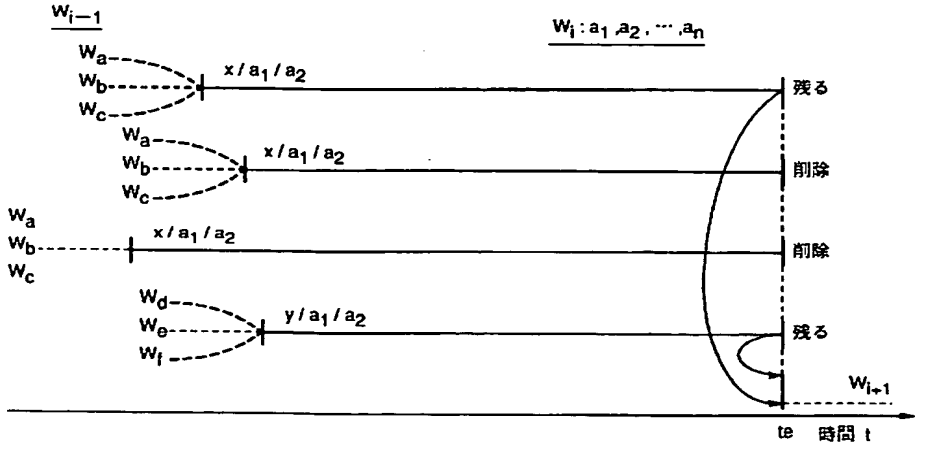
【図１】



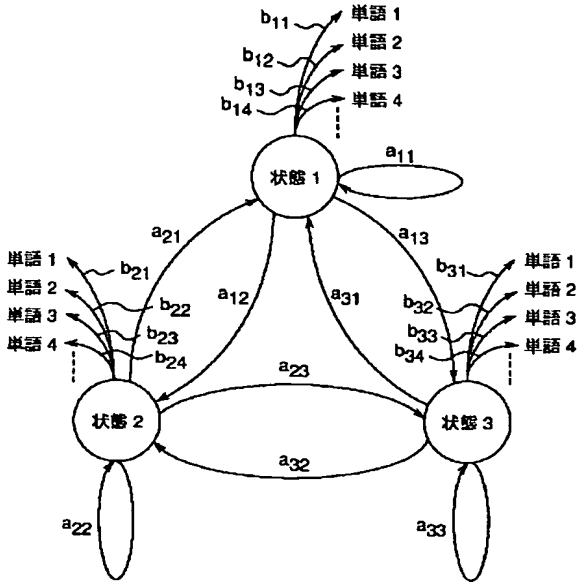
【図２】



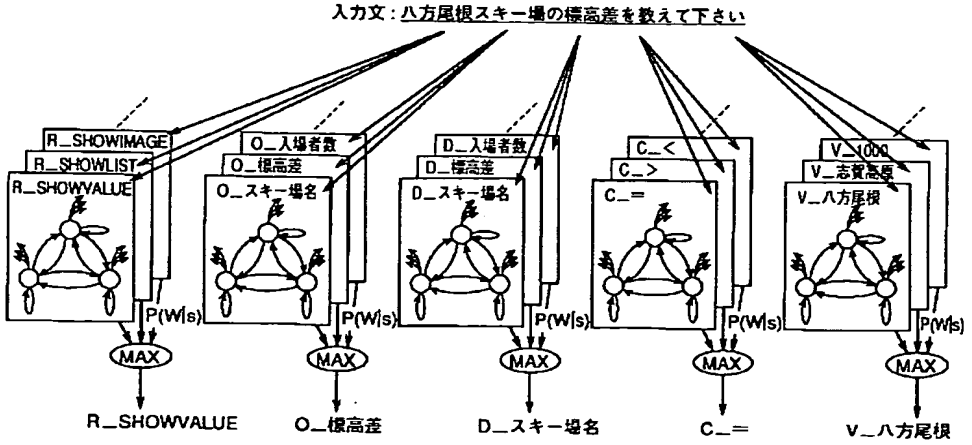
【図３】



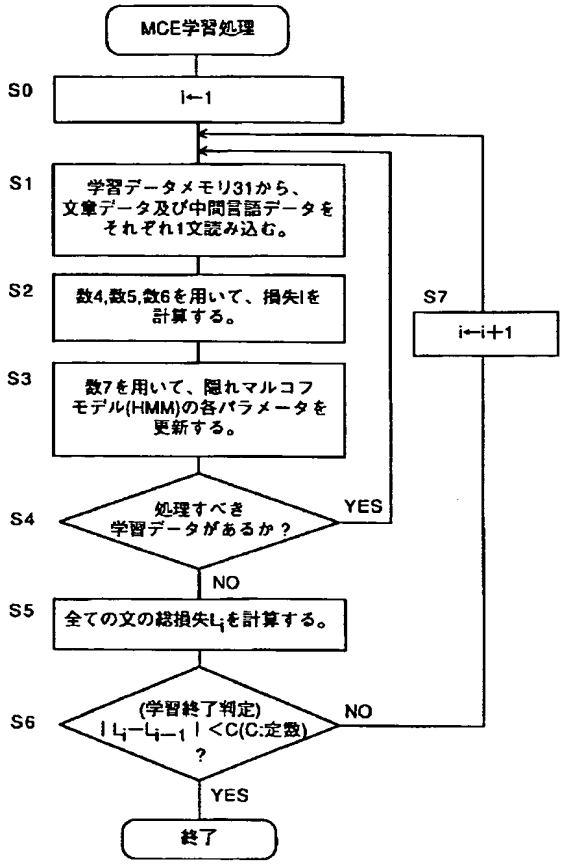
【図４】


状態 1
状態 2
状態 3
単語 1
単語 2
単語 3
単語 4
b_{11}
b_{12}
b_{13}
b_{14}
a_{11}
a_{21}
a_{12}
a_{13}
a_{31}
a_{23}
a_{32}
a_{22}
a_{33}
b_{21}
b_{22}
b_{23}
b_{24}
b_{31}
b_{32}
b_{33}
b_{34}

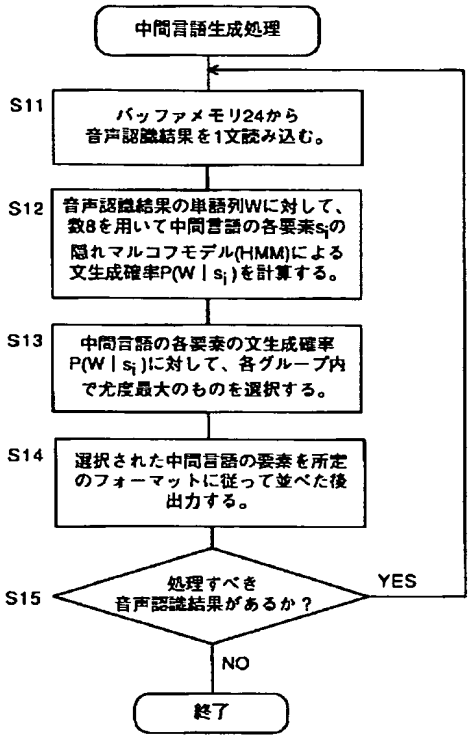
【図５】


入力文：八方尾根スキー場の標高差を教えて下さい
R_SHOWIMAGE
R_SHOWLIST
R_SHOWVALUE
O_入場者数
O_標高差
O_スキー場名
D_入場者数
D_標高差
D_スキー場名
C_<
C_>
C_=
V_1000
V_志賀高原
V_八方尾根
P(W|s)
MAX
R_SHOWVALUE
O_標高差
D_スキー場名
C_=
V_八方尾根

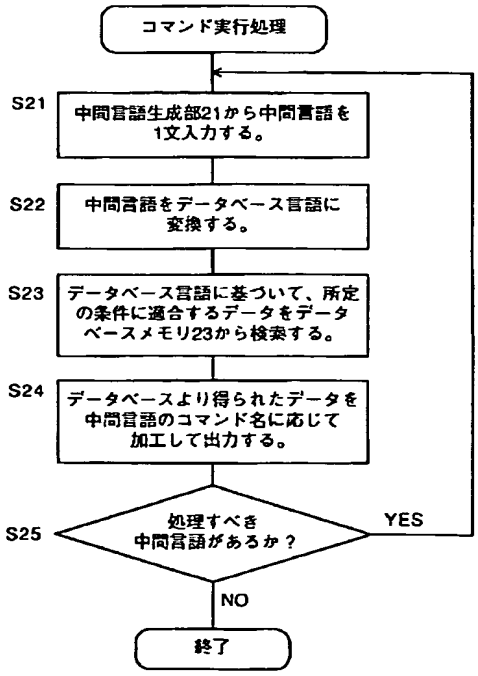
【図６】


MCE学習処理
S0
i←1
S1
学習データメモリ31から、文章データ及び中間言語データをそれぞれ1文読み込む。
S2
数4,数5,数6を用いて、損失lを計算する。
S3
数7を用いて、隠れマルコフモデル(HMM)の各パラメータを更新する。
S4
処理すべき学習データがあるか？
YES
NO
S5
全ての文の総損失L_iを計算する。
S6
(学習終了判定)
$|L_i-L_{i-1}|<C$(C:定数)
？
NO
YES
S7
i←i+1
終了

【図７】


中間言語生成処理
S11
バッファメモリ24から音声認識結果を1文読み込む。
S12
音声認識結果の単語列Wに対して、数8を用いて中間言語の各要素s_iの隠れマルコフモデル(HMM)による文生成確率$P(W|s_i)$を計算する。
S13
中間言語の各要素の文生成確率$P(W|s_i)$に対して、各グループ内で尤度最大のものを選択する。
S14
選択された中間言語の要素を所定のフォーマットに従って並べた後出力する。
S15
処理すべき音声認識結果があるか？
YES
NO
終了

【図８】


コマンド実行処理
S21
中間言語生成部21から中間言語を1文入力する。
S22
中間言語をデータベース言語に変換する。
S23
データベース言語に基づいて、所定の条件に適合するデータをデータベースメモリ23から検索する。
S24
データベースより得られたデータを中間言語のコマンド名に応じて加工して出力する。
S25
処理すべき中間言語があるか？
YES
NO
終了

フロントページの続き

(51)Int.Cl.[6]		識別記号	FI		
Ｇ０６Ｆ	17/30		Ｇ０６Ｆ	15/38	Ｐ
// Ｇ０６Ｆ	3/16	３２０		15/403	３１０Ｚ